Stratification of meta-analyses based on risk of bias is appropriate and does not induce selection bias.

Matthew J Page[1], Asbjørn Hróbjartsson[2,3,4], Camilla Hansen[2,3,4], Jonathan AC Sterne[5,6]

1. School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
2. Centre for Evidence-Based Medicine Odense (CEBMO), Odense University Hospital, Odense, Denmark
3. OPEN – Open Patient data Explorative Network, Odense University Hospital, Odense, Denmark
4. Department of Clinical Research, University of Southern Denmark, Odense, Denmark
5. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
6. NIHR Bristol Biomedical Research Centre, Bristol, UK

Correspondence to: Dr. Matthew Page, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, Victoria, 3004, Australia. Telephone: +61 9903 0248. Email address: matthew.page@monash.edu

Word count: 919

Keywords: Bias; Methodology; Quality; Randomized trials; Systematic reviews

We read, “Stratification by quality induces selection bias in a meta-analysis of clinical trials” by Stone et al. (1) with interest. We think that the authors have misunderstood the implications of their findings, and that their conclusions could mislead authors of systematic reviews.

Stone et al. argue that restriction of meta-analyses to lower risk of bias (higher quality) studies, or stratification of meta-analyses according to risk of bias (or quality) leads to selection bias (also known as collider bias). Stone et al. repeated the analyses originally reported by Jüni et al. in 1999 (2) by applying 25 different quality assessment tools to the 17 clinical trials included in the low molecular weight heparin meta-analysis originally reported by Nurmohamed et al. (3). They followed Jüni et al. in stratifying the trials based on a quality threshold, and finding that the extent of discrepancies between intervention effect estimates in trials classified as ‘high’ and ‘low’ quality varied according to the quality scale used. They also observed small-study effects – a tendency for effects estimated in smaller studies to differ from those estimated in larger studies (4) – more often in the lower-quality strata than in the higher-quality strata and attributed these results to the stratification (conditioning) on quality. They concluded that discrepancies between the pooled estimates in the high- and low-quality trials “are largely because of conditioning on quality, which allows precision and effect size to associate within strata by quality” (1).

We believe that Stone et al. have misunderstood the implications of their proposed directed acyclic graph (DAG), which is shown in Panel A of Figure 1. Under the assumptions encoded in the DAG, a lower quality ranking is a common effect of ‘reasons for biased studies’ and ‘reasons for smaller studies’. Stratification on quality ranking will therefore induce an association between study precision and study results within quality strata. However, it does not follow that the results within strata defined by quality ranking are biased. This would only be the case if results in smaller studies were biased as well as less precise, which is not assumed in the DAG.

Figure 1. Directed acyclic graphs (DAG) depicting assumed relationships between bias and study size, as proposed by Stone et al. (1) (Panel A) and with "methodological expertise" as a common cause of bias and study size (Panel B), and study size additionally a cause of bias (Panel C).

We believe there are other factors linking study size with risk of bias that Stone et al. have not considered. In the more plausible DAG shown in Panel B of Figure 1, greater methodological expertise is a common cause of larger study size and lower risk of bias. In this case, conditioning on risk of bias will remove the association between biased results and less precise results. Alternatively, the DAG in Panel C of Figure 1 below represents a situation in which both risk of bias and study size influence results, because smaller studies with statistically non-significant effects are less likely to be published. In this case, it would be necessary to condition on both risk of bias and study size.

"Quality" and "risk of bias" are very different concepts, though Stone et al. do not distinguish between them clearly. Quality of a trial may include many aspects that are not directly relevant for internal validity, such as type of patients included or sample size. Risk of bias, in contrast, focusses only on bias understood as internal validity. This is reflected in the fundamental differences between "quality scales" and tools to assess "risk of bias". Neither sample size nor whether a power calculation was reported have direct implications for risk of bias in the result an individual study, which is why they are not included in either version 1 (5) or 2 (6) of the Cochrane risk of bias tool for randomized trials. However, quality scales often include these factors. Stone et al. calculated a composite quality score incorporating all items across the 25 scales, one of which posed the question, "Was there ≥50 patients per group?" In addition, 11 of the 25 scales included the question, "Was there a sample size justification before study and power declared a priori?" In these cases, lower quality score is an effect of sample size or whether a power calculation was reported. It is

therefore unsurprising that in their case study, Stone et al. found precision and effect size to associate within strata by quality.

Assessing risk of bias targets the key question of whether results of studies included in a systematic review should be believed, by focusing only on factors likely to influence estimated intervention effects. By contrast, quality scales often include items that are not directly related to bias, such as "Are both inclusion and exclusion criteria mentioned?", "Was the therapeutic regimen adequately described for the treatment and control group", and "Were the criteria for measuring outcomes clearly described?" (1). For this reason, we discourage the use of quality scales (including the composite scale developed by Stone et al.) to appraise included studies, and recommend authors focus on risk of bias.

Based on their findings, Stone et al. recommend "The common strategy of assessing the impact of quality in meta-analysis by excluding lower or including higher quality studies should be abandoned" (1). However, a single old case study is not a good basis for drawing general recommendations for conduct of modern systematic reviews. Restriction or stratification according to risk of bias should remain the recommended approach for incorporating risk of bias assessments in meta-analyses.

Competing Interests

We have read the journal's policy and the authors of this manuscript have the following competing interests: MJP and AH are co-convenors of the Cochrane Bias Methods Group, and CH is coordinator of the Group. The views expressed in this manuscript are those of the authors and not necessarily those of Cochrane or its registered entities, committees, or working groups.

Funding

There was no direct funding for this manuscript. MJP is supported by an Australian National Health and Medical Research Council (NHMRC) Early Career Fellowship (1088535). JACS is supported by a National Institute for Health Research Senior Investigator award NF-SI-0611-10168. The funders had no role in decision to publish or preparation of the manuscript.

Author Contributions

All authors declare that they meet the ICMJE conditions for authorship. MJP wrote the first draft of the manuscript. All authors contributed to revisions of the manuscript. All authors approved the final version of the submitted manuscript.

References

1. Stone J, Gurunathan U, Glass K, Munn Z, Tugwell P, Doi SAR. Stratification by quality induced selection bias in a meta-analysis of clinical trials. Journal of clinical epidemiology. 2019;107:51-9.
2. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. Jama. 1999;282(11):1054-60.
3. Nurmohamed MT, Rosendaal FR, Buller HR, Dekker E, Hommes DW, Vandenbroucke JP, et al. Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. Lancet (London, England). 1992;340(8812):152-6.
4. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. Journal of clinical epidemiology. 2000;53(11):1119-29.
5. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.

6. Higgins JPT, Sterne JAC, Savović J, Page MJ, Hróbjartsson A, Boutron I, et al. A revised tool for assessing risk of bias in randomized trials. In: Chandler J, McKenzie J, Boutron I, Welch V (editors). Cochrane Methods. Cochrane Database of Systematic Reviews. 2016;Issue 10(Suppl 1):29-31.

Figure 1. Directed acyclic graphs (DAG) depicting assumed relationships between bias and study size, as proposed by Stone et al. (1) (Panel A) and with "methodological expertise" as a common cause of bias and study size (Panel B), and study size additionally a cause of bias (Panel C).